0.1 Multilayer Feedforward Networks are Universal Approximators

In 1988 Maxwell Stinchcombe and Halber White show that Multilayer Feedforward Networks with one hidden layer are Universal Approximators.

The proof is based on that arbitrary squashing function are capable of approximating any Borel measurable function from one finite dimensional space to another to any desired degree of accuracy, provided sufficiently hidden units are available. In this sense Feedforward networks are a class of Universal approximators. (REWRITE THREE LAST SENTENCES)

Firstly we are going to start with some basic definitions. Particularly we are going to introduce familiar of function from more particular to more general that will represent

Definition 0.1.1. Affine functions

Affine functions are defined as all the function of the form $A(x) = w \cdot x + b$ from $\mathbb{R}^d$ to $\mathbb{R}$ where $d \in \mathbb{N} - \{0\}$, w and x are vector in $\mathbb{R}^d$, $\cdot$ denotes the usual dot product of vector, and $b \in \mathbb{R}$ is a scalar.

Definition 0.1.2. Squashing function

A function $\psi : \mathbb{R} \longrightarrow [0, 1]$ is a squashing function if

- It is non-decreasing.
- $\lim_{x\to\infty} \psi(x) =$.
- $\lim_{x\to-\infty} \psi(x) = 0$.

Squashing functions are measurable due to the fact that have at most countably discontinuities.

Most common examples are:

- Threshold functions.[1]
- Indicator functions: $\psi(\lambda) = 1_{\{\lambda>0\}}$.
- The ramp function: $\psi(\lambda) = \lambda 1_{\{0\leq\lambda\leq1\}} + 1_{\{\lambda>1\}}$
- The cosine squasher of Gallant and White (1988)

$$\psi(lambda) = (1 + cos(\lambda + 3\frac{\pi}{2})\frac{1}{2})1_{\{\frac{-\pi}{2}\leq\lambda\leq\frac{\pi}{2}\}}1_{\{\frac{\pi}{2}<\lambda\}}$$

Definition 0.1.3. General family for NN

In order to introduce the familiar class of function for single hidden layer feedforward networks. Now we are going to define the familiar class of function of output functions

[1] A threshold function is a function that takes the value 1 is a specified function of the arguments exceeds a given threshold and 0 otherwise. **[threshold-definition]**

For any Borel measurable function $\mathcal{G}(\cdot)$ mapping $\mathbb{R}$ to $\mathbb{R}$ and $r \in \mathbb{N} - \{0\}$, let define the class of function

$$\sum^{r}(\mathcal{G}) = \{f : \mathbb{R}^r \longrightarrow \mathbb{R} =$$

$$f(x) = \sum_{j=1}^{q}(\beta_j \mathcal{G}(A_j(x)), x \in \mathbb{R}^r, \beta_j \in \mathbb{R}, A_j \in A^r, l_j, q \in \mathbb{N} - \{0\})\}$$

When $\mathcal{G}$ is a squashing function, $\sum^r(\mathcal{G})$ is the familiar class of we were searching for, the familiar class of output function for single hidden layer feedforward networks at the output layer.

The scalars β_j correspond to network weight from hidden to output layers.

Definition 0.1.4. More general family definition

For any measurable function $\mathcal{G}(\cdot)$ mapping $\mathbb{R}$ to $\mathbb{R}$ and $r\mathbb{N} - \{0\}$, let

$$\sum\prod^{r}(\mathcal{G}) = \{f : \mathbb{R}^r \longrightarrow \mathbb{R}|$$

$$f(x) = \sum_{j=1}^{q}, \beta_j \cdot \prod_{k=1}^{l_j} \mathcal{G}(A_{jk}(x)), x \in \mathbb{R}^r, \beta_j \in \mathbb{R}, A_{jk} \in A^r, l_j, q \in \mathbb{N} - \{0\})\}$$

Definition 0.1.5. Multivariable continuos functions and Borel σ-field

Let $\mathcal{C}^r$ be the set of continuos functions from $\mathbb{R}^r$ to $\mathbb{R}$, and let $\mathcal{M}^r$ be the set of all Borel measurable functions from $\mathbb{R}^r$ to $\mathbb{R}$. We denote the Borel σ-field of $\mathbb{R}^r$ as $\mathcal{B}^r$

In order to write a self-contained book we are going to define some basic analysis and algebra concepts.

Definition 0.1.6. Ring [**rudin-lebesgue-theory**] A family $\mathcal{R}$ of sets is called a ring if $A \in \mathcal{R}$ and $B \in \mathcal{R}$ implies

1. $A \cup B \in \mathcal{R}$.
2. $A \cap B \in \mathcal{R}$. TODO ADD SINCE 301

Definition 0.1.7. $\sigma-$ring [**rudin-lebesgue-theory**]

A ring $\mathcal{R}$ is called a $\sigma-$ ring if

1. $\cup_{n=1}^{\infty} A_n \in \mathcal{R}$ whenever $A_n \mathcal{R} for n \in \mathbb{N} - \{0\}$.

TODO ADD SINCE PAG 301

TODO

Definition 0.1.8. Additive functions [**rudin-lebesgue-theory**] We define a additive function ϕ, as a set function defined on $\mathcal{R}$ if ϕ assigns to every $A \in \mathcal{R}$ a number $\phi(A) \in \mathbb{R}$ and if for every $A, B \in mathcalR$ that verified $A \cap B = \emptyset$, this implies

$$\phi(A \cup B) = \phi(A) + \phi(B),$$

In addition, ϕ is defined as *countably additive* if $A_i \cap A_j = \emptyset (i \neq j)$ implies

$$\phi(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \phi(A_n)$$

We assume that the range of ϕ does not contain both $+\infty$ and $-\infty$. Also we exclude set functions whose only value is $+\infty$ and $-\infty$.

When ϕ is countably additive and $\phi(A) \geq 0$ for every $A \in \mathcal{R}$, It is called a *non negative countably additive function.*

Definition 0.1.9. Measure spaces [**rudin-lebesgue-theory**]

The notation that we are going to use is

Definition 0.1.10.

- Let $\mathcal{C}^r$ be the set of continuous functions from $\mathbb{R}^r$ to $\mathbb{R}$.
- Let $\mathcal{M}^r$ be the set of all Borel measurable functions from $\mathbb{R}^r$ to $\mathbb{R}$.
- The Borel $\sigma-$ field of $\mathbb{R}^r$ is going to be denoted as $\mathcal{B}^r$

The classes $\sum^r(\mathcal{G})$ and $\sum\prod^r(G)$ belong to $\mathcal{M}^r$ for any Borel measurable $\mathcal{G}$. When $\mathcal{G}$ is continuos,

TODO pag 3-8